Weekly Report

15th Jan – 4th Feb 2018

To 20th Jan

Done:

1. The paper revision is almost done.

2. Michael has done a very basic framework of the system. I read his code, fix a small bug, and continue learning some skills; Annie implemented some basic vis of data with uncertainty.

3. Since Michael is not available in the days to come, I have asked Prof. Chen to allocate 3 undergraduate students in China. However, they are also not available at present since they are preparing their final exams.

4. Discussion with Annie:

(a) Use some uncertainty vis method to develop a contour of the detection data (point data, with confidence);

(b) We may use the radiative power data and uncertainty methods to estimate the fire influenced areas. It is kind of a prediction, based on satellite detection data with different granularities, and radiative power. We know there is a lag of fire data reporting, and we may use such a method to do some quick response, like based on the detection data, what's the probability a certain place will be influenced by the fire? Probability here may be derived from the uncertainty propagation, based on Annie's opinion.

(c) we may try to combine perimeter data with detection data, to apply some uncertainty techniques.

(d) Correlate data from the different source is still a problem. If we don't do preprocessing, we need to do time-space query. For weather, we may find the nearest station(s); for reports, based on latitude, longitude, time, place name; for detection data, need time, geo-operation in mongodb. Reports and perimeter data are the cores; all other data center on them

(e) currently, we may consider more options for different views, especially the summary view and the map. for the guys in management office to choose.

To do:

1. Although I'm still struggling with codings, I will begin coding the time series data parts. Also, figure out how to transform Annie's code to the code can be run in our framework.
2. determine the strategy of correlate/fuse the data.
3. discuss more about the uncertainty vis part of the system.

To 27th Jan

This week, I focused on coding and debugging, and tried to organize the ideas.

Next week, I will continue coding. Besides, I will collect papers on ML simple approaches to predict the probability of wildfire.
Currently, as we discussed this week and previously, the prediction/recommendation idea in our systems are:
(1) Annie's uncertainty vis itself can provide some predict information;
(2) use the meteorological, fuel, and smoke data *before* the fire to train models. The model should be simple and interactive (for potential users to use and understand, like regression, decision tree, etc.).
(3) As we considered before, there are labs that focus on precise fire propagation/prediction, and that's difficult for us to compete with their resource and technique. But we may try to combine Annie's uncertainty vis result and the model trained in (2) to give a coarse estimation of fire propagation (i.e., *during* the fire).

During this weekend, I will try to read several papers and refine the data processing task that the undergraduate student will do. the tasks for the undergraduate student:

To test some basic models, we need to divide fire areas into grids. We need information about these grids (like weathers, smokes, and fuels). Thus, for example, we need to query the weather station which is nearest to one grid to get the weather condition at one timestamp.

This website provides APIs for weather query https://synopticlabs.org/api/mesonet/. Based on the fire area, we need the weather info which is close to each grid. The fire perimeters are stored here:https://drive.google.com/drive/folders/1qsSID0bEXTYOxrYzITjjvPy2SYH4CcSZ?usp=sharing

Not every fire has a detailed report, only those big ones do. For example, in this page(2017 Cal Fire, please select "Fires per page: all" to view all the fire incidents) http://cdfdata.fire.ca.gov/incidents/incidents_archived, only part of the incidents "Name" row have hyperlinks. Click on one hyperlink will forward to another page which shows detailed information. Some fires have 'reports' links in the "Condition" row. Please help us crawl these info.

Smoke records are here: http://www.ssd.noaa.gov/PS/FIRE/kml.html, but it only preserves certain days of data.

Fuel data can be calculated based on this, http://www.southernfireexchange.org/SFE_Publications/factsheets/2016-1.pdf. However, it's still not enough as model input. I am trying to find more reliable sources, and if you can find more please tell me.

Besides, we are trying to organize these data, since data are from different sources with different temporal/spatial granularities. How to organize them for better query/analysis performance.

To 4th Feb

Data:
Look at the data and find many issues
1 . CalFire website provides fire data, but not all of them are wildfire; InciWeb website has more detailed description of fire characters / fire causes, but seems to be subset of CALFIRE records
2. Some small fires (burned area is small) may not be detected. (may be detected by one source/multiple source/or none)
3. small fires with short duration may not be suitable to train model, but we still have to figure out the boundary between "small" and "big" fire
4. find one fire record, in CALFIRE is 14Oct, in perimeter data is 23Oct, and 1km detection data has no record, 4km record is placed at an adjacent place to perimeter data.
5. for Mesonet weather API, not every station has same sensors
All in all, align these data/ data cleaning need much effort.

Model:
1. find some papers which help calculate fire danger indicators
2. learn more about how other papers divide fire area into grids /
3. find several fire models which are simple, and papers regarding interactive vis of those simple models (regression models)

Code:
1. made some CSS modification, will do more at the weekend (basic skills of interaction and animation, which combine D3 and Reactjs)

Other:
1. Finish revising of a rejected paper

To do:
1. Data cleaning, pick out several fire datasets to do model test
2. add weather data into the system, add basic interactions

I'm still reading papers, which use some interactive regression methods/ and interpretable regression/decision/ensemble learning methods.